DOCUMENT RESUME

ED 198 172 TM 810 140

AUTHOR Livingston, Samuel A.
TITLE Nonverbal Communication Tests as Predictors of Success in Psychology and Counseling.
INSTITUTION Educational Testing Service, Princeton, N.J. Center for Occupational and Professional Assessment.
REPORT NO ETS-COPA-80-01
PUB DATE Jun 80
NOTE 38p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Correlation; *Counseling Effectiveness; Higher Education; Interpersonal Competence; *Nonverbal Communication; *Nonverbal Tests; *Predictive Measurement; *Predictive Validity; *Psychology; Scores; Student Evaluation
IDENTIFIERS Group Embedded Figures Test (Altman et al); Inter Person Perception Test; Photo Classification Test (Livingston et al); Photo Sequence Comprehension Test; Profile of Nonverbal Sensitivity (Rosenthal et al); Test of Implied Meanings (Sundberg)

ABSTRACT

The selection and development of six tests measuring the ability to receive and interpret nonverbal communications are described, as is an attempt to gather evidence of their value as predictors of success in two occupations requiring high levels of interpersonal skills-- psychology and counseling. The tests were: (1) Inter-Person Perception Test; (2) Test of Implied Meanings; (3) "PONS-audio from the Profile of Nonverbal Sensitivity; (4) "PONS-visual" from the Profile of Nonverbal Sensitivity; (5) Photo Sequence Comprehension Test; and (6) Photo Classification Test (PCT). The 106 subjects were graduate students and had taken also the Group Embedded Figures Test (GEFT) and the Graduate Record Examination (GRE) or the Miller Analogies Test. The principal criterion measure was a faculty rating form used to rate the student on a nine-point scale in: academic work, interpersonal relations with peers, faculty, and clients; personal qualities; and predicted effectiveness as a psychologist or counselor. Ratings were collected several months after testing. The PCT seems to have been the most effective predictor, but it was uncorrelated with the GRE and only weakly correlated with the GEFT, as were most of the other nonverbal communication tests. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED198172

COPA-80-01

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

NONVERBAL COMMUNICATION TESTS AS PREDICTORS OF SUCCESS IN PSYCHOLOGY AND COUNSELING

Samuel A. Livingston

Center for Occupational and
Professional Assessment
Educational Testing Service

June, 1980

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. C. Widenmiller

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

COPA

TM 810 140

Copyright © 1980, Educational Testing Service. All rights reserved.

ABSTRACT

Six tests of nonverbal communication skills were investigated in an attempt to improve prediction of success for psychologists and counselors. The subjects were graduate students at two different schools; the criterion variables were faculty members' judgments of the students' academic work, interpersonal relations, personal characteristics, and "predicted effectiveness" in the profession. Faculty ratings were collected several months after students were tested. One of the six nonverbal communication tests predicted faculty ratings of several characteristics at both schools. This test was uncorrelated with the Graduate Record Examinations and only weakly correlated with the Group Embedded Figures Test, as were most of the other nonverbal communication tests.

ACKNOWLEDGMENT

I would like to thank Peter Nathan, Ruth Schulman, and Aaron Nierenberg, their faculty colleagues, and the students who participated in the study, for making this research possible. I would also like to express my appreciation for the technical assistance of Sylvan Tomkins and for the advice and encouragement of Robert Rosenthal, Judith Hall, Donald Goodenough, Philip Oltman, and Herman Witkin.

NONVERBAL COMMUNICATION TESTS
AS PREDICTORS OF SUCCESS IN
PSYCHOLOGY AND COUNSELING

This project grew out of an attempt to improve the prediction of professional success by measuring skills that are not measured by traditional aptitude tests. Most traditional aptitude tests focus on verbal and mathematical skills, and, to a lesser extent, spatial and abstract reasoning skills. While these skills are important in many professions, there is another group of skills that are also of great importance in many professions: interpersonal skills.

The difficulties of measuring interpersonal skills in any reasonably objective and efficient way are enormous. Interpersonal behavior is an interactive process consisting of complex responses to complex stimuli. The stimuli are difficult to standardize and the responses are difficult to evaluate. One promising approach to the problem is that of role-playing simulations (Levine and McGuire, 1968, 1970), but these measures are expensive to administer and score. Even if they were to become much more widely used than they have been, there would still be a need for measures that can be administered to several examinees at once and scored by machines or by persons without special training.

Most, if not all, interpersonal activities involve some form of communication. It seems reasonable to assume that a person cannot be effective at an occupation requiring a high level of interpersonal skill unless that person can communciate effectively. Communication in our culture usually involves the use of words, but it almost always includes many nonverbal behaviors as well. Tests of verbal skills are widely used in occupational selection and counseling. Tests of nonverbal communications skills are not.

Communications skills can be divided into two types: sending and receiving. In the case of nonverbal communication, the receiving skills are much easier to measure than the sending skills; the encoded messages can be standardized in the form of pictures and recordings, and the examinee's responses can be obtained in a multiple-choice format. For this reason, most existing tests of nonverbal communications skills are tests of receiving skills, rather than sending skills.

Some research in this area has been done by Rosenthal and his colleagues (Rosenthal, et al, 1974, 1979) using their Profile of Nonverbal Sensitivity (PONS), which yields separate subscores for various combinations of face, body, and voice communications. They found that skill at interpreting nonverbal communication, as measured by the PONS, tended to increase with age up to the late teens. "Impaired groups"--psychiatric patients and alcoholic patients--did not perform as well as normal adults. More relevant to the present study were the differences between occupational groups. Actors tended to do quite well, but clinical psychologists performed only about as well on the PONS as college students, and teachers and business executives performed no better, on the average, than high school students. One attempt to use a portion of the PONS to predict occupational success was reported by McClelland and Fiske (1974). They used the twenty random-spliced audio items as a separate test, with an adjustment based on the test-taker's age.* The

*The adjustment added to the test-taker's score one percent of the number of questions missed for every year of age over 35. Thus, a person of 45 who answered 12 of the 20 questions correctly would receive an adjusted score of 12 plus 10 percent of 8, or 12.8.

subjects were human service workers in a large Eastern city. The criterion variable was a supervisor's rating of the worker as "outstanding" or "average"; 26 of the 51 subjects were rated as "outstanding." The age-adjusted scores on the random-spliced PONS items correlated .45 with the supervisors' ratings. Other correlational studies cited by Rosenthal, et al, (1979, p. 372) reported positive correlations between total scores on the full PONS (220 items) and supervisors' ratings of the professional skills of teachers and of clinicians.

Is it reasonable to expect tests of nonverbal receiving skills to be useful as indicators of interpersonal skills in general? The answer to this question is yes, if either of the following two conditions is true:

1. Much of the variation between persons in interpersonal effectiveness is directly caused by differences in ability to receive and interpret nonverbal messages, or
2. Nonverbal receiving skills are highly correlated with other interpersonal skills.

If either of these conditions is true, tests of the ability to receive and interpret nonverbal communications could be valuable instruments in the selection, guidance, and placement of persons into or out of occupations requiring high levels of interpersonal skills--particularly if the tests are objectively scored and suitable for group administration. The project described in this report involved the selection and development of several such measures and an attempt to gather evidence of their value as predictors of success in two such occupations.

METHOD

Test selection and development

There are at least four important ways in which people transmit nonverbal messages in situations associated with their work: tone of voice, facial expression, body position, and the use of interpersonal space. The tests investigated as part of this project involved all of these four channels of communication, either separately or in combination with each other.

The first phase of this project consisted of the selection of tests that seemed promising as predictors of success in a wide range of occupations involving interpersonal skills, including the modification of one existing test and the development of two new tests. In reviewing tests purporting to measure interpersonal skills, the investigator excluded tests that appeared to depend heavily on verbal skills, tests based on stimulus materials that appeared to be highly contrived or artificial (e.g., cartoons or obviously posed photographs), and tests that required the examinee to have knowledge, training, or experience in a particular occupation. The tests selected for inclusion in the study were the Inter-Person Perception Test (Heussenstamm and Hoepfner, 1969), the Test of Implied Meanings (Sundberg, n.d.) and two tests from the Profile of Nonverbal Sensitivity (Rosenthal, et al, 1974, 1979).

The Inter-Person Perception Test (IPPT) is a purely nonverbal test that focuses entirely on the face. Each item consists of a photo of a person's face, followed by four photos of a second person's face. The examinee is instructed to "mark the letter for one of the four faces on the right that is thinking or feeling the same as the first person." Both persons in an item are

of the same race and sex. The test contains forty items, five for each race/sex combination of both sexes and four races: White (Euro-American), Black (Afro-American), Japanese (Japanese-American), and Chicano (Mexican-American). This test was used without modification.

The Test of Implied Meanings is a test intended to measure a person's ability to perceive and interpret the messages conveyed by inflection and tone of voice in spoken language. The test is administered by tape recording. Each item consists of a short sentence or portion of a sentence, spoken in such a way as to imply more than the words themselves indicate. The examinee must choose the one of four written phrases that expresses the additional meaning implied by the way the sentence is spoken. The original version of this test consisted of forty items, all read by narrators. The version used in this study consisted of twenty items from the original test, plus ten items created from tape recordings of "natural" speech in meetings and discussion groups. In constructing these ten newly created items, the investigator determined the correct implied meanings from portions of the tape-recorded discussion that were not included in the brief excerpt played for the examinee.

The Profile of Nonverbal Sensitivity (PONS) consists of several different types of items. Each item is formed from a videotape recording of a young woman, simply and neatly dressed, role-playing a common situation. The test taker must choose the one of two brief phrases that correctly describes what the woman is doing (e.g., "expressing gratitude"). The original PONS consists of 220 items presented in movie form. The version used in this study consists of two 40-item tests. One test ("PONS-audio") is administered by tape recording. Each audio item is based on a two-second fragment of the audio portion of the recording, with the words made unintelligible. In half the items the tape

recording has been "content filtered" to blur the speech by removing the high frequencies. In the other half of the items the tape has been subjected to "random splicing", i.e., the tape is cut into very short fragments, which are randomly shuffled and reassembled, producing a garbled version of the original speech excerpt. The second test ("PONS-visual") is based on still photographs made from the PONS videotape. Half the photos show only the woman's face; the other half show her body from the shoulders to the knees. She is dressed identically in all the photos. The process of development of the PONS is described in detail by Rosenthal, et al (1979).

Two additional tests were developed for this project by the investigator with the aid of a consultant, an expert in the interpretaion of nonverbal communication.* Both those tests are based on photographs, but, unlike those in the IPPT and the PONS, the photos in these tests all show an interaction involving two or more persons.

The Photo Sequence Comprehension Test is similar to a reading comprehension test in format. However, instead of reading a paragraph, the examinee looks at a sequence of 10 or 12 photographs of two or more people in a situation involving some degree of interpersonal interaction (e.g.. a business discussion in an office). Each sequence of photos is followed by four multiple-choice questions about the actions, thoughts, and feelings of the people pictured.

To develop the Photo Sequence Comprehension Test the investigator first obtained the permission of the persons in the photos to photograph them while they engaged in their normal workday activities. (Exception: one of the five sequences was role-played, although the persons involved were all acting in

---

*Dr. Sylvan Tomkins

roles they were accustomed to.) The investigator took a series of approximately 60 photos, using available light to avoid the intrusion of a flash. After the photos were printed, the investigator arranged them in sequence for the consultant, who "read" the nonverbal messages they contained. The investigator then summarized the verbal content of the interaction for the consultant, and the consultant selected a sequence of 10 or 12 photos in which the nonverbal messages told the story of what had happened. The investigator and consultant then decided what questions to ask about the photo sequence. This process was repeated for each of the five sequences on the test, with four questions based on each sequence.

The Photo Classification Test is an entirely nonverbal test. Each item-set consists of nine photos showing the same two people in the same setting (e.g., a man handing some papers to a woman sitting behind a typewriter). On the left page are six photos, classified into two criterion groups labeled A and B. On the right page are three more photos. The examinee's task is to classify each of these three additional photos as belonging with group A or group B. The examinee is not told the basis for the classification, but must infer it by looking at the photos in groups A and B. In fact, the test is constructed so that the feelings expressed by the people in the pictures form the basis for the classification. Other aspects of the situation (e.g., the placement of physical objects) vary in ways that do not consistently distinguish between group A and group B.

The Photo Classification Test was developed by having the same persons role-play a common situation several times. Two different versions of each situation were role-played. The persons and the setting were the same in

both versions, but the "script", i.e., the instructions given orally to the role-players, was changed, so as to imply a different emotional content in the two versions. (Exception: one of the five item-sets was made up of photos taken during an actual business meeting and selected with the aid of the consultant who assisted in the development of the Photo Sequence Comprehension Test.)

Subjects

The subjects for the study were students in graduate programs in psychology and counseling at two different institutions. School 1 is a graduate school located at a major state university. The school has Ph.D and Psy.D. programs in clinical psychology and in school psychology. Most of the students are full-time graduate students. School 2 is a small college which has a master's-degree program in academic counseling. Classes in this program meet in the evening, and many of the students are not full-time students. Participation at both schools was voluntary, and the students were not paid. Participation consisted simply of taking the tests and signing a form allowing the investigator to have access to the student's academic records and to discuss the student's professional work with faculty members. At School 1 the students were released early from a regular class to take the tests. At School 2 the tests were given during a regular class period.

Instruments

In addition to the nonverbal communication tests, the students participating in the study took the Group Embedded Figures Test (Oltman et al, 1971). This test is made up of problems that require the test-taker to locate a simple

geometric figure that is embedded in a complex geometric pattern. The test consists of a practice set of seven problems, followed by two separately timed sets of nine problems each.

Additional predictors were available to the investigator for many of the students at School 1, where many of the students had taken either the Graduate Record Examination or the Miller Analogies Test. The students' scores on these tests were included in the data analysis.

The principal criterion measure for the study was a faculty rating form. This form was a modified, expanded version of a portion of an existing form used for practicum supervisors' evaluations at School 1. The form called for the faculty member to rate the student on a nine-point scale in each of several categories: academic work; interpersonal relations with peers, faculty, and clients; several personal qualities (clarity, discretion, enthusiasm, initiative, organization, poise, responsibility, sensitivity, tact, warmth); and, finally, "predicted effectiveness" as a psychologist or counselor. On the rating form each variable was identified only by a single word or brief phrase; no additional descriptive information was provided. In addition, the rating scale for each specific variable contained an extra space labeled "cannot rate student". The schools were asked to distribute rating forms for each participting student to all faculty members who had taught the student. Faculty members were asked to rate only those students they felt they knew well enough to rate meaningfully. Faculty ratings were obtained for 66 students at School 1 and 40 students at School 2. The number of faculty members rating each student at School 1 varied from one to twelve. At School 2 each student was rated by two faculty members.

Procedure

The tests were administered on several occasions. At School 1 there were testing sessions in February and March of 1977, September of 1977, and September

of 1978. At School 2 there were testing sessions in October of 1977 and September of 1978. Faculty ratings were collected at both schools in the spring of 1978 and 1979. The time interval between a student's testing session and the collection of faculty ratings for that student varied from nine months to 21 months. None of the faculty members at either school had access to any information about the students' performance on the tests.

The tests were administered to groups ranging in size from seven to approximately thirty students. The investigator began by asking the students to read a description of the research study and to sign a statement of informed consent. The description of the research study described the tests as being "based on pictures and sound recordings" and "intended to measure skills that are not measured by conventional academic aptitude tests." The first test administered was the Group Embedded Figures Test. It was followed by the Photo Classification Test, the Test of Implied Meanings, the Photo Sequence Comprehension Test, the PONS-audio, the PONS-visual, and the Inter-Person Perception Test, in that order.

Because the students were not all rated by the same faculty members, the effects of the raters' differing standards had to be removed from the criterion data. This step was accomplished by a regression procedure which assumes that each individual rating of a student by a teacher can be expressed as the sum of a student-quality parameter, a teacher-leniency parameter, and a random variable. In mathematical notation, let x_{ij} represent the rating of student i by teacher j. Then

$$x_{ij} = a_i + b_j + e_{ij}$$

where a_i is the student-quality parameter for student i, b_j is the teacher-leniency parameter for teacher j, and e_{ij} is the remainder, assumed to be the

result of random variation. The a_i parameter is the average of the ratings that student i would have received if all students had been rated by all teachers. The regression procedure estimates this parameter for each student. The procedure was done separately for the two schools and separately for each variable on the rating form. The estimated student-quality parameters (the a_i in the model) were then used as criterion measures. Wherever the term "faculty ratings" appears in the Results section of this report, it refers to these estimated student-quality parameters.

RESULTS

The key question to be answered by the data is whether or not any of the tests effectively predicted the judgments that the faculty members made several months later, after working closely with the students. This question can best be answered by an examination of the correlations between the test scores and the faculty ratings. The sample size for these correlations varies somewhat, because some faculty members did not rate all of the students on all of the variables and because a few students did not take all of the tests. The sample sizes for the correlations range from 57 to 66 at School 1 and from 35 to 40 at School 2. For "predicted effectiveness," which was intended as the main criterion measure, the sample sizes range from 64 to 66 at School 1 and from 38 to 40 at School 2. (The minimum correlations necessary for statistical significance at the five percent level are therefore approximately .24 at School 1 and .33 at School 2.)

Table 1 shows the correlations of the tests with the faculty ratings. An inspection of these correlations reveals some interesting results. First, the correlations for School 2 are generally larger than those for School 1. This fact may be a consequence of differences in the reliability of the ratings (discussed below). In particular, the ratings of academic work at School 2 seem to have been predicted quite well by three of the nonverbal communication tests.

The Photo Classification Test seems to have been the most effective predictor. In fact, for many of the ratings, including "predicted effectiveness," it was the only effective predictor. In Table 1 there are only five combinations of predictor and criterion for which the correlation is at least .20 at both schools. In four of these five combinations, the predictor is the Photo Classification Test.

Table 1. Correlations of tests with faculty ratings. (Upper numbers refer to School 1, lower numbers refer to School 2.)

	Academic Work	Interpersonal Relations with ...				Clarity	Discretion	Enthusiasm
		supervisors	patients/ students	faculty	peers			
Inter-Person	.06	*	-.08	-.15	-.14	.06	.09	-.08
Perception Test	.41	-.14	-.15	-.21	-.18	.36	-.16	-.01
Test of Implied	-.10	*	.18	.20	.33	.06	.05	-.11
Meanings	.04	.06	-.07	.21	.09	.41	.01	.05
Profile of Nonverbal	.01	*	-.01	.03	.05	-.09	.06	.07
Sensitivity—audio	.17	-.26	-.20	-.22	-.16	.05	-.01	-.09
Profile of Nonverbal	.02	*	.06	-.10	-.09	-.07	-.13	-.02
Sensitivity—visual	.38	.25	-.04	.44	.24	.25	.22	.04
Photo Sequence	.01	*	-.10	.03	-.02	.26	.01	-.08
Comprehension Test	.07	-.34	-.16	-.41	-.32	.08	-.40	-.18
Photo Classification	.21	*	.19	.16	.10	.24	.07	.33
Test	.45	.02	-.16	.08	-.17	.37	.15	.01
Group Embedded	.17	*	-.12	.01	.03	.05	-.04	-.12
Figures Test	.33	-.05	-.23	.02	-.20	.24	.14	-.28

*This variable was rated at school 2 only.

Table 1 (cont.)

	Initiative	Organization	Poise	Responsibility	Sensitivity	Tact	Warmth	Predicted Effectiveness
Inter-Person Perception Test	-.11 .04	-.01 .16	-.03 .04	-.05 .19	.00 -.03	.01 -.23	-.13 -.23	.03 -.04
Test of Implied Meanings	.05 .00	-.03 .27	.12 .21	-.04 .20	.03 .27	.05 .22	.07 .07	.07 .18
Profile of Nonverbal Sensitivity--audio	.10 -.12	.11 -.08	-.07 -.03	.11 -.21	-.06 -.03	.02 -.09	-.07 -.10	.00 -.10
Profile of Nonverbal Sensitivity--visual	-.01 -.08	.01 .37	-.13 .21	-.10 .34	-.10 .10	-.03 .32	-.03 .16	-.08 .14
Photo Sequence Comprehension Test	-.11 -.06	-.06 -.13	-.02 -.08	-.10 -.10	.00 -.23	-.09 -.26	-.03 -.31	.07 -.20
Photo Classification Test	.22 .25	.20 .30	.05 .28	.14 .36	.12 .23	.03 .11	.01 -.04	.24 .26
Group Embedded Figures Test	.06 -.22	.10 .17	-.07 -.12	-.02 .16	-.10 -.25	-.07 -.01	-.15 -.30	.04 .07

The PONS-visual and, to a lesser extent, the Test of Implied Meanings seem to have predicted several of the ratings effectively at School 2, but not at School 1. The PONS-audio and the Photo Sequence Comprehension Test failed to predict any of the ratings (except possibly in the negative direction, i.e., with higher scorers on these tests receiving poorer ratings).

The Group Embedded Figures Test predicted some of the faculty ratings, particularly at School 2. The test correlates positively with ratings of academic work, "clarity," and "organization"; negatively with ratings of interpersonal relations with patients or students, interpersonal relations with peers, "sensitivity," and "warmth."

Table 2 shows the range of possible scores, the mean, standard deviation, and reliability coefficient* for each test, and the intercorrelations among the tests. The data in Table 2 are based on larger samples than those in Table 1, because they include some additional students who took the tests but were not rated by any faculty members. The sample sizes range from 80 to 83 at School 1 and from 48 to 50 at School 2. Obviously, the tests do not correlate highly with each other; the highest correlation between any two tests is .39, and only five of the 42 correlations between tests are larger than .20 in size. One surprising finding is the very low reliability of the PONS tests, especially in light of the success of the visual PONS in predicting several of the faculty ratings at School 2.

---

*For the two PONS tests these are KR-20 (or "alpha") coefficients. For the other tests they are split-halves or, in the case of the Photo Classification Test, split-thirds coefficients, corrected with the Spearman-Brown formula. Attempts to use the split-halves method with the PONS tests yielded correlations less than zero in two of the four cases, even though the splits were made for maximum similarity between the two half-tests.

Table 2. Test score data. (Upper numbers refer to School 1; lower numbers refer to School 2.)

	IPPT	Implied Meanings	PONS-audio	PONS-visual	Photo Sequence	Photo Classification	Embedded Figures
Possible score range	0-40	0-30	0-40	0-40	0-20	0-15	0-18
Chance score	10	7.5	20	20	5	7.5	*
Mean	25.1	19.8	25.8	29.8	12.4	9.2	12.4
	23.4	18.9	26.0	29.8	11.6	8.0	9.6
Standard deviation	3.3	3.9	3.4	2.7	2.5	3.1	4.6
	3.6	3.8	2.6	2.7	2.2	3.0	5.3
Reliability coefficient	.35	.68	.46	.19	.39	.85	.83
	.40	.75	.07	.28	.37	.74	.87
Intercorrelation with							
Inter-Person Perception Test		-.02	-.08	-.05	.04	-.13	.02
		.16	.28	.06	.39	.12	.02
Test of Implied Meanings			-.04	.08	.09	-.12	.32
			.13	.25	.09	.29	.22
PONS-audio				-.07	.05	.11	-.05
				.03	.11	.16	.11
PONS-visual					-.02	-.21	-.13
					-.18	-.09	-.10
Photo Sequence Comprehension						.02	.33
						.18	.11
Photo Classification							.10
							.20

*The Group Embedded Figures Test is not a multiple-choice test.

The scores on most of the tests were slightly higher at School 1 than at School 2. The data do not suggest any "floor" or "ceiling" effects. Only the Photo Classification Test and the Group Embedded Figures Test were "speeded" in the conventional sense, i.e., so that a slow-working examinee would not reach the last items on the test. This speededness probably accounts for the fact that several students scored below chance on the Photo Classification Test.* Two interesting features of the scores on the Group Embedded Figures Test are not reflected in the data in Table 2. The score distribution was very flat or possibly bimodal, and there was a noticeable practice effect between the two separately timed halves of the test.

Table 3 shows the inter-rater reliability of the faculty ratings. The numbers in Table 3 represent the proportion of the variation in the ratings that is accounted for by differences between students. Almost without exception, the reliability of the ratings is higher at School 2 than at School 1. This difference in reliability of the ratings may account for the fact that the correlations between test scores and faculty ratings tended to be larger at School 2 than at School 1. The most reliable ratings were the ratings of academic work at School 2, and these were the ratings best predicted by the nonverbal communication tests. Most of the reliabilities at School 1 were in the range of .40 to .50. Using these reliability estimates to "correct for unreliability in the ratings" (i.e., to estimate the correlation between the test and a perfectly reliable rating) would increase the reported correlations for School 1 by aproximately 40 to 60 percent. For example, the correlation between the Photo Classification Test and the "predicted effectiveness" rating would increase from .24 to .37.

---

*It may also account for the surprisingly high reliability of the Photo Classification Test, which is probably partly spurious, since the split-thirds were not separately timed.

Table 3. Inter-rater reliability of faculty ratings.

	School 1	School 2
Academic work	.52	.78
Interpersonal relations with		
supervisors	*	.54
patients/students,	.25	.47
faculty	.36	.62
peers	.56	.55
Clarity	.45	.57
Discretion	.43	.65
Enthusiasm	.40	.61
Initiative	.40	.58
Organization	.51	.59
Poise	.37	.48
Responsibiity	.47	.55
Sensitivity	.40	.53
Tact	.41	.56
Warmth	.27	.47
Predicted effectiveness	.42	.60

*Rated only at School 2.

Table 4 shows the intercorrelations of the faculty ratings. The correlations tend to be positive, with only a few exceptions. The highest correlations are where they might be expected, e.g., between "academic work" and "clarity," between "enthusiasm" and "initiative," and between "organization" and "responsibility." Similarly, the lowest correlations are where they might be expected, e.g., between "academic work" and "warmth." The faculty at School 1 seem to have placed a greater emphasis on intellectual qualities in rating "predicted effectiveness" than did the faculty at School 2, as indicated by the correlations in the last row of Table 4. However, "predicted effectiveness" correlates at least .45 with all the other ratings at both schools and thus seems to fulfill its intended purpose as a single overall criterion measure. A word of caution may be in order, though: the rating most highly correlated with "predicted effectiveness" is "interpersonal relations with faculty." Nevertheless, it seems reasonable to use "predicted effectiveness" as an overall criterion measure wherever a single measure is necessary to avoid burdening the reader with huge quantities of statistical information.

Several of the tests are constructed in such a way that meaningful subscores can be defined. The Photo Sequence Comprehension Test and the Photo Classification Test each consist of five separate sets of items, with each item set based on a single collection of pictures. In the Inter-Person Perception Test, the items can be classified according to the race or sex of the persons pictured. The Test of Implied Meanings can be subdivided into three groups of items on the basis of the speakers in the recording; only the items in the third group are based on excerpts of "natural" conversation. The audio-PONS can be subdivided into "content-filtered" and "random-spliced" speech excerpts. The visual-PONS can be subdivided into pictures showing the face only and those

Table 4. Intercorrelations of faculty ratings.
(Upper numbers are for School 1;
lower numbers, for School 2.)

	Academic Work	Interpersonal -supervisors	Interpersonal -patients/students	Interpersonal -faculty	Interpersonal -peers	Clarity	Discretion	Enthusiam	Intiative	Organization	Poise	Responsibility	Sensitivity	Tact	Warmth
Interpersonal -supervisors	*														
	.30														
Interpersonal -patients/students	.05	*													
	.07	.55													
Interpersonal -faculty	.32	*	.71												
	.37	.79	.45												
Interpersonal -peers	.31	*	.58	.70											
	-.02	.67	.68	.67											
Clarity	.61	*	.49	.57	.38										
	.83	.34	.12	.38	-.01										
Discretion	.37	*	.53	.64	.46	.57									
	.11	.69	.25	.65	.56	.08									
Enthusiasm	.58	*	.62	.65	.37	.54	.37								
	.23	.53	.54	.40	.34	.26	.10								
Initiative	.62	*	.59	.61	.23	.39	.47	.77							
	.37	.37	.35	.29	.12	.33	.00	.84							
Organization	.78	*	.37	.45	.14	.57	.64	.57	.84						
	.77	.50	.08	.51	.07	.65	.22	.41	.53						
Poise	.31	*	.54	.55	.45	.57	.67	.37	.40	.46					
	.41	.49	.37	.51	.45	.42	.32	.31	.35	.36					
Responsibility	.60	*	.55	.59	.34	.54	.68	.64	.77	.84	.59				
	.77	.60	.23	.54	.16	.70	.31	.48	.62	.88	.51				
Sensitivity	.33	*	.64	.51	.60	.52	.57	.39	.32	.39	.72	.59			
	.12	.49	.56	.52	.48	.21	.43	.37	.27	.10	.34	.22			
Tact	.34	*	.59	.62	.50	.45	.84	.33	.43	.55	.71	.59	.69		
	.15	.67	.51	.84	.72	.21	.71	.28	.12	.17	.48	.23	.71		
Warmth	.09	*	.49	.55	.68	.21	.43	.41	.26	.12	.47	.27	.58	.58	
	-.07	.68	.74	.56	.77	-.07	.46	.46	.20	.04	.34	.17	.64	.64	
Predicted effectiveness	.61	*	.69	.76	.48	.71	.69	.73	.67	.65	.66	.70	.55	.66	.45
	.47	.76	.70	.74	.62	.47	.51	.60	.51	.46	.54	.61	.61	.71	.69

*Rated only at School 2.

showing the body only. Both PONS tests can be divided into subscores on the basis of the similarity of the correct and incorrect options, in terms of the feeling expressed--positive or negative--and the status of the speaker in reaction to the person spoken to--dominant or subordinate. (See Rosenthal, et al., 1979, pp. 29-30.)

Table 5 presents the correlations of these subscores with the faculty ratings and also the mean percent-correct for each subtest, at each of the two schools. For the Inter-Person Perception Test and the Test of Implied Meanings the analysis did not reveal any substantial differences between subscores in difficulty or any high correlations of subscores with predicted effectiveness. In the audio-PONS, the analysis revealed only that the four items in which the two answer options were similar in both ways (positivity of feelings expressed and status of the speaker) were harder, in general, than the other items. In the visual-PONS, the body items were somewhat more difficult than the face items, and the items in which the two answer options were similar in status of the speaker were slightly harder than those in which the answer options differed in this respect. None of the subscores on either PONS test correlated substantially with predicted effectiveness. However, at School 2 the face items subscore on the visual PONS correlated .34 with ratings of academic work; the correlation for the body items subscore was .17. The analysis of subscores on the Photo Sequence Comprehension Test also revealed some differences in the difficulty of the different item sets, but no subtest correlated with predicted effectiveness at both schools.

The analysis of subscores on the Photo Classification Test yielded some interesting results. The second and third item sets were much easier than the others and, unlike the others, did not correlate well with predicted

Table 5. Subtests: Mean percent correct and correlation with predicted effectiveness.

Test and subtest	No. of items	Mean percent correct		Correlation with predicted effectiveness	
		School 1	School 2	School 1	School 2
Inter-Person Perception Test					
Ethnicity of person pictured:					
White	10	59	60	-.05	.07
Black	10	58	63	.06	-.22
Chicano	10	58	56	-.11	-.01
Japanese	10	64	62	.14	.12
Sex of person pictured					
Male	20	.62	.62	.03	.06
Female	20	.63	.59	.01	-.13
Test of Implied Meanings					
Items 1-10	10	64	61	.14	.10
Items 11-20	10	69	67	.07	.06
Items 21-30	10	63	61	-.10	.26
PONS-audio					
Method of construction:					
Random-spliced	20	65	66	.00	-.08
Content-filtered	20	66	64	-.05	-.08
Similarity of options:					
Positivity and dominance	4	46	43	.07	-.20
Positivity only	10	66	64	.07	-.33
Dominance only	13	66	70	-.07	.13
Neither	13	66	68	-.03	.14
PONS-visual					
Picture content					
Face	20	78	80	-.15	.18
Body	20	70	67	.03	.01
Similarity of options:					
Positivity and dominance	10	72	72	-.09	.12
Positivity only	10	75	75	-.18	.16
Dominance only	6	73	70	-.02	.07
Neither	14	77	76	.08	-.02
Photo Sequence Comprehension Test					
Item set 1	4	71	72	-.08	-.27
Item set 2	4	51	49	-.02	.06
Item set 3	4	86	82	-.04	-.02
Item set 4	4	69	62	.09	.05
Item set 5	4	50	38	.23	-.05
Photo Classification Test*					
Item set 1	3	72	58	.30	.44
Item set 2	3	90	80	.05	-.15
Item set 3	3	93	83	-.12	.09
Item set 4	3	68	45	.41	.54
Item set 5	3	57	46	.20	.17

*The correlations involving subtests of this test do not include students who left all items in the subtest blank.

effectiveness. The first and fourth item sets correlated strongly with predicted effectiveness at both schools. Because of the speededness of this test, the correlations involving the fourth and fifth item sets were based on a self-selected subsample of the students that excluded the slower workers; persons who left all items in an item set blank were excluded in computing the correlation. Therefore, the correlations reported in Table 5 for these item sets may be somewhat smaller than the correlations that would have been observed for these item sets if they had been placed earlier in the test.

One question of considerable practical importance concerns the relationship between the nonverbal communication tests and commonly used academic aptitude tests: To what extent do the tests investigated in the present study provide information that is not provided by the more commonly used tests? Some of the students participating in the present study had taken the Graduate Record Examinations (GRE), the Miller Analogies Test, or both. At School 1, scores were available for 42 students on the GRE verbal and quantitative tests, for 48 students on the GRE advanced test in psychology, and for 15 students on the Miller Analogies Test. The correlations between these tests, the nonverbal communication tests, and the faculty ratings of "academic work" and "predicted effectiveness" are shown in Table 6. The correlations among the three GRE scores are .60 (verbal with quantitative), .52 (verbal with advanced psychology), and .41 (quantitative with advanced psychology). Because of the small number of students taking the Miller Analogies, the correlations involving this test should be regarded only as suggestive. No correlations were computed for School 2, where only 8 students had taken the GRE and only 8 had taken the Miller Analogies.

Table 6. Correlations with commonly used academic predictors. (School 1 only)

	Graduate Record Examinations			
	Verbal (n=42)	Quantitative (n=42)	Advanced Psychology (n=48)	Miller Analogies (n=15)
Tests:				
Inter-Person Perception Test	.19	.07	.04	-.29
Test of Implied Meanings	.37	.33	.07	.52
PONS-audio	.08	.00	.21	-.22
PONS-visual	-.05	-.22	-.15	.35
Photo Sequence Comprehension Test	.11	.03	.18	.11
Photo Classification Test	.05	-.05	.06	.03
Group Embedded Figures Test	.38	.49	.16	.27
Faculty ratings:				
"Academic work"	.31	.24	.59	.49
"Predicted effectiveness"	.16	.11	.39	.56

The correlations in Table 6 indicate that in general, the nonverbal communication tests showed quite low correlations with the more conventional predictors. The Photo Classification Test, which was the only effective predictor at School 1, correlated essentially zero with all three GRE scores (and with Miller Analogies also). The visual PONS actually correlated negatively with all three GRE scores. Incidentally, the GRE advanced psychology score appears to be an effective predictor of faculty ratings at School 1 (and the Miller Analogies Test may also be an effective predictor). The GRE verbal and quantitative scores do not appear to be nearly as effective, even though they correlate fairly well with the GRE advanced psychology score.

DISCUSSION

The results of the present study are mostly in the form of correlations computed from samples ranging in size from approximately 40 persons to approximately 80 persons. With data of this type, we must beware of over-interpretation, i.e., attempting to explain "results" that would not generalize to future samples or to the population we are interested in. However, when the same two variables correlate approximately .25 in two completely independent samples, we are probably justified in concluding that the relationship between them is real. And if the two independent samples represent different populations of students, faculty raters, and institutional settings, we have some grounds for expecting the result to generalize to other, similar populations.

Why, then, was the Photo Classification Test the only one of the nonverbal communication tests to correlate with faculty ratings at both schools? At least two important features distinguish this test from the other tests used in the study. First, the test-taker's task is considerably more complex. Instead of simply choosing one picture or descriptive phrase from two or four options presented, the test taker must look at six highly similar photos that have been classified into two groups and determine the basis for the classification. Only then can the test-taker correctly classify the remaining three photos. Second, the photos in this test were "stage directed" so that features that were irrelevant to the classification were deliberately varied in ways that did not consistently distinguish between the two groups of photos. That is, features of the photos that had nothing to do with nonverbal communication or with the feelings role-played by the persons in the photos were varied, but not in ways that would permit the test-taker to distinguish the Group A photos from the Group B photos on that basis.

The analysis of subscores from the Photo Classification Test provides some evidence for the importance of these irrelevant features in the photos. Two of the five item-sets were especially effective as predictors of the faculty ratings. An inspection of the test reveals that in both these item sets (and not in the other three), one of the irrelevant features that was deliberately varied occupies a prominent and conspicuous place in the photos. This fact suggests that the effectiveness of the test may be in identifying persons who can perceive and interpret nonverbal messages in the presence of distracting stimuli.

The instructions used for the Photo Classification Test in this study did not inform the students that the basis for the classification had anything to do with nonverbal communication or with feelings and emotions. The sample item used to introduce the classification task was chosen so as not to help the students discover this fact. The photos in the sample item simply showed a man at work in an office. In the "Group A" pictures he was seated at the desk; in the "Group B" pictures he was standing behind the desk. The effectiveness of the test may depend on the students' not knowing in advance that the test involves nonverbal communications. If so, the test would be best used for guidance, rather than selection, so that test-takers would have no reason to attempt to prepare for the test. In either case (guidance or selection), the information gained from this test would be an addition to the information provided by commonly used academic aptitude tests; there is practically no overlap.

While the Photo Classification Test showed positive correlations with most of the faculty ratings at one or both of the two schools, it predicted some

of the ratings much better than others. It seems to have been most effective at predicting "academic work," "clarity," "initiative," "organization," and "predicted effectiveness." It seems to have been least effective at predicting interpersonal relations with persons other than faculty, "tact," and "warmth." This pattern of results suggests that the test may function as an indicator of the extent to which the test-taker is inclined and able to view interpersonal situations analytically. (The test does not seem to be simply a test of general analytic ability; its correlations with the Group Embedded Figures Test and the Graduate Record Examinations were quite low.)

The lack of positive results for the other five nonverbal communication tests is surprising and disappointing. Four of those tests were selected because they appeared to the author to be especially promising; the fifth was the author's own creation. The lack of success for the random-spliced subtest of the audio-PONS is particularly surprising in light of the positive results reported by McClelland and Fiske (1974). The visual-PONS was correlated with several of the faculty ratings at School 2 but not at School 1; it may be worth investigating further.

The pattern of correlations between the Group Embedded Figures Test and the faculty ratings at School 2 reflects the pattern one would expect on the basis of the theory of field-dependence (Witkin, et al., 1977). High scorers on this test tended to be rated higher in academic work, "clarity," and "organization" than low scorers, but lower in interpersonal relations with peers, "sensitivity," and "warmth." Witkin, et al. (1977) stated that ". . . there has not been a real check on the expected relation between field dependence and better performance in educational domains where a social orientation is emphasized." The

results of this study suggest that the nature of this relationship will depend heavily on how "better performance" is defined. If it is defined in terms of academic work, the field-independent students, who score high on Embedded Figures, are likely to do better than the field-dependent students. If it is defined in terms of such personal characteristics as "enthusiasm," "sensitivity," and "warmth," the field-dependent students are likely to do better.

REFERENCES

Heussenstamm, F. K. and Hoepfner, R. Inter-Person Perception Test, Form AA. Monitor, 1969.

Levine, H. G., and McGuire, C. Role-playing as an evaluative technique. Journal of Educational Measurement, 1968, 5, 1-8.

Levine, H. G. and McGuire, C. H. The use of role-playing to evaluate affective skills in medicine. Journal of Medical Education, 1970, 45, 700-705.

McClelland, D. C. and Fiske, S. T. Validation of a human service worker test (revised edition). Boston: McBer and Company, 1974.

Oltman, P. K., Raskin, E., and Witkin, H. A. Group Embedded Figures Test. Palo Alto, California: Consulting Psychologists Press, 1971.

Rosenthal, R., Archer, D., DiMatteo, M. R., Koivumaki, J. H., and Rogers, P. L. The language without words. Psychology Today, September 1974, 8(4), 64-68.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., and Archer, D. Sensitivity to Nonverbal Communication. Baltimore: Johns Hopkins University Press, 1979.

Sundberg, N. Test of Implied Meanings. Unpublished, n.d.

Witkin, H. A., Moore, C. A., Goodenough, D. R., and Cox, P. W. Field-dependent and field-independent cognitive styles and their educational implications. Review of Educational Research, 1977, 47(1), 1-64.